From actions to suggestions: supporting the work of biologists through laboratory notebooks

Marcello Sarini*, Enrico Blanzieri**, Paolo Giorgini**, Claudio Moser***
*Dipartimento di Informatica, Sistemistica e Comunicazione
Università degli Studi di Milano-Bicocca
Via Bicocca degli Arcimboldi, 8 – 20126 MILANO-ITALY
**Dipartimento di Informatica e Telecomunicazioni
Università di Trento
Via Sommarive, 14 -38050, Povo, TRENTO-ITALY
***Istituto Agrario San Michele all'Adige,
Via E. Mach, 1-38010 San Michele a/Adige (TN)-ITALY
email: sarini@disco.unimib.it, {blanzier, pgiorgio}@dit.unitn.it, claudio.moser@mail.ismaa.it

Abstract. The paper presents an approach and a technology to support the work of biologists in their laboratories. By means of analysis of related literature, informal discussions with people working in laboratory, and experience of one the authors with laboratory work in genomic research, we identified different biologists' information needs and different strategies in order to satisfy them. A relevant part of biologists' work practice is about the continuous annotation of their work in their personal notebooks. We claim that observing notebooks can boost a technology which is a generalization of recommendation systems and provides biologists with suggestions concerning their work. These suggestions are generated by combining the observation of biologists performed actions in the past with actions patterns incorporating biologists' work practices that are discovered by the system out of the transcriptions of biologists' notebooks.

1 Introduction

Computer Supported Cooperative Work (CSCW) is the art of providing meaningful mediations between sociality of people and technological support in cooperative work. One of the main CSCW issues is to understand what collaboration is about. There are various forms of collaboration (e.g., coordination, use of common shared resources) and different social mechanisms people developed in order to let collaboration happen, such as artifacts [1], protocols [2], awareness of information [3], definition of conventions [4], and negotiation of meanings [5, 6]. However, depending on the situation, these mechanisms may be not effective enough and possibly technologies may play a supporting role [7].

Traditionally, Knowledge Management (KM) focuses on knowledge as something that can be separated from the people who produced it. This view overlooks the central role of people interactions that enable members of communities, ranging from small groups to large organizations, to create and exchange different forms of knowledge. Again, the use of these forms of knowledge can be supported by proper technologies, when needed [8].

In both areas CSCW and KM, the introduction and use of technologies taking into account the consolidated work practices is a major concern.

The aim of our work is to provide a technology that supports people in performing their work and acknowledges the value of their work practices. Following an approach developed in Artificial Intelligence [9] and applied to Information Retrieval [10, 11], the proposed technology generalizes the concepts of Recommendation Systems [12] to arbitrary complex scenarios where users receive suggestions concerning their work. The level of support provided by the system is limited by the possibility of observing users actions and by the capability of the system in generating meaningful suggestions. In this view, the proposed technology deals with two critical issues. First, to collect the work practices related to a certain domain in order to make them shared by the members of a community; second, to orient the workers towards the fulfillment of their goals, still leaving them free to behave according to their own work practices.

In this paper the technology is applied to support the work of biologists in their laboratories. We think this field could greatly benefit from this technology due to the tension between the often implicit nature of biologists' work practices and biologists' heavy information needs. In fact, in order to properly carry out their work, biologists have to apply specific protocols, handle reagents and instruments but often the related practices are just in the scientists' minds and rarely are shared among them. An important source of information in biological laboratories is the laboratory notebook where biologists register the actions they perform during their work. In Nonaka's terms [13], this behavior leads to the *externalization* of scientists' tacit knowledge concerning their work practices. The system collects single scientist work practices, combines these pieces of knowledge converting it into patterns of actions performed by the scientists and makes them available to all of them as a mean to satisfy their current information needs. The information provided by the system then helps biologists *internalize* what they experienced, thus enriching their tacit knowledge.

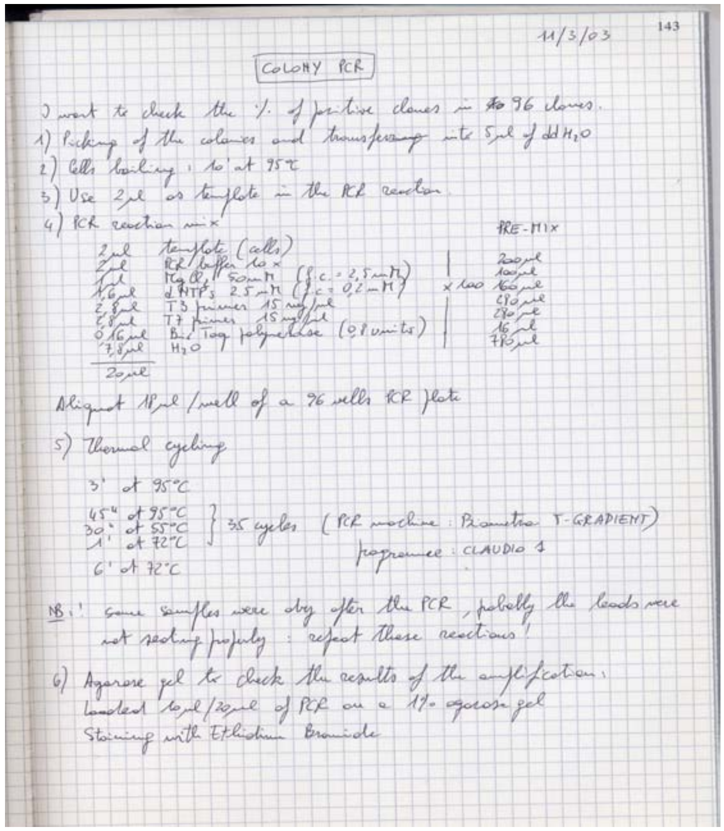
The paper is organized as follows. The next two sections describe biological work inside laboratories. In particular the first one focuses on the nature of biological work, while the second one identifies biologists information needs in order to complete their work and discusses how these needs may be satisfied. The remaining sections present the solution we propose in order to support biologists in accomplishing their work.

2 The nature of biological work in the laboratories

Let us now consider the nature of biological work inside laboratories. The following description is the result of direct interactions with people involved in university laboratories, the experience of one the authors with laboratory work in genomic research, and the reading of related literature (see for instance [14]).

During their work, scientists manage materials and instruments in order to conduct experiments. In managing materials they have to find them when needed: this is not always easy since materials can be in use by others laboratory members at the same time or can have been stored in an unknown location. A good laboratory practice is also to check in advance if there are enough materials for running a specific experiment. The use of laboratory instruments is often complex because it requires experience and their ad-hoc tuning. Therefore, experiments have to be planned in advance combining the information about the needed material, for the instruments tuning, and also for the procedure to follow. A typical example is represented by a PCR (Polymerase Chain Reaction) experiment. This is a very common enzymatic reaction that is used in biological and medical research in order to amplify a specific DNA fragment.

A fundamental resource supporting laboratory work is the scientist's notebook. An individual notebook contains the record of everything a scientist does [15].



11/3/03 143

COLONY PCR

I want to check the % of positive clones in the 96 clones.
1) Picking of the colonies and transferring into 5µl of dd H_2O
2) Cells boiling: 10' at 95°C
3) Use 2µl as template in the PCR reaction
4) PCR reaction mix — PRE-MIX

2µl	template (cells)	
2µl	PCR buffer 10x	200µl
1µl	$MgCl_2$ 50mM (f.c. = 2,5mM)	100µl
1,6µl	dNTPs 2,5mM (f.c. = 0,2mM) x100	160µl
2,8µl	T3 primer 15 ng/µl	280µl
2,8µl	T7 primer 15 ng/µl	280µl
0,16µl	Bio Taq polymerase (0,8 units)	16µl
7,8µl	H_2O	780µl
20µl		

Aliquot 18µl/well of a 96 wells PCR plate

5) Thermal cycling
3' at 95°C
45" at 95°C, 30" at 55°C, 1' at 72°C } 35 cycles (PCR machine: Biometra T-GRADIENT) programme: CLAUDIO 1
6' at 72°C

NB: some samples were dry after the PCR, probably the leads were not sealing properly: repeat those reactions!

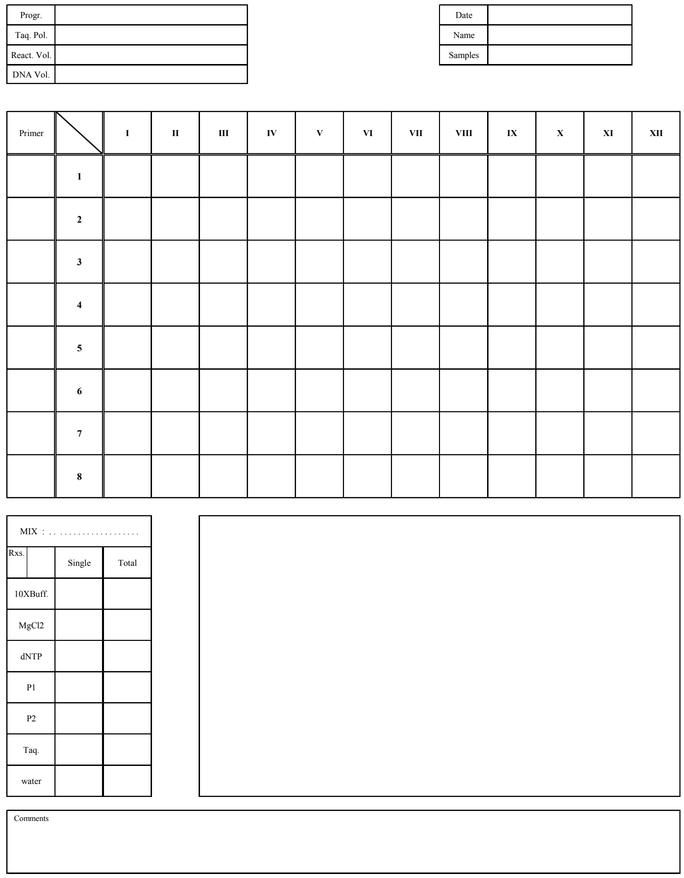
6) Agarose gel to check the results of the amplification:
loaded 10µl/20µl of PCR on a 1% agarose gel
Staining with Ethidium Bromide

Figure 1: an example of a notebook page

Figure 1 shows an example of a laboratory notebook page reporting on a PCR experiment using a bacterial lysate (colony PCR). In the first row the main objective of the experiment is stated. Items 1 to 3 explain the procedure to prepare the bacterial lysate and specify how much of it has to be used in the PCR reaction. Item 4 lists all the components and the volume to be added in the PCR reaction mix specifying the concentration of each stock. Item 5 summarizes temperatures and times that will be used during the thermal cycling. This is the step where the amplification occurs. In item 5 is also reported an observation written after the thermal cycling was finished ("some samples were dry…") and a perspective for future work ("repeat those reactions"). Item 6 very briefly describes the analytical procedure used for checking the results of the PCR experiment.

Usually, scientists write on their notebooks the steps to go through in a detailed manner before to start the experiment which often is repeated several times introducing slight changes from one time to the other. A typical case is a PCR experiment performed for the development of molecular markers used in the construction of genetic linkage maps: several couples of primers are tested on the same or few types of genomic DNA templates. This implies that only the primers will vary among reactions while all the other reagents and also the actions to be performed will not change. According to this observation, the biologists designed a form (see Figure 2) which is an artifact reflecting the structure emerged in performing PCR experiments and which is used to organize biologists' work accordingly. This form, printed on a paper sheet, is filled according to the specific PCR experiment; it is then used like a recipe during the setting-up of the reactions and finally is attached in the lab-notebook.

Although notebooks are very personal they are consulted as a relevant support for the strict group of people who share the same interests. This is particularly true when an experiment is carried out involving different laboratory members in different times. In this case notebooks play the role of coordination artifacts [1] among the involved people, since they objectify the current state of the experiment and suggest the new actions to perform, according to the scientist's experience.



Progr.	
Taq. Pol.	
React. Vol.	
DNA Vol.	

Date	
Name	
Samples	

Primer		I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
	1												
	2												
	3												
	4												
	5												
	6												
	7												
	8												

MIX :		
Rxs.	Single	Total
10XBuff.		
MgCl2		
dNTP		
P1		
P2		
Taq.		
water		

Comments

Figure 2: a form used by biologists for PCR experiments

3 Information needs in the laboratory scenario

The competence of one of the author in laboratory work allowed us to define a more precise scenario by focusing on scientist's information needs either to successfully complete her work or to understand why some performed experiments failed. These information needs are related to different aspects of ordinary working practices: *physical*, *scientific*, and *social*.

Physical aspects are about the use of materials which have to be found by the scientist while working in the laboratory. Different materials are stored in different places and it may be not easy for a scientist to find them at the right moment. As an example, consider that some reagents have to be stored in a freezer with a temperature of -80° C.

Scientific aspects are related to the knowledgeable practices of scientists in conducting their work: for instance, about successful and unsuccessful experiments, the proper tuning of instruments for specific experiments, and so on.

Social aspects concern the direct and indirect collaboration between people working in the laboratory. An example are the agreed upon conventions for keeping the laboratory notebooks making them easily consulted by other people.

Let us now consider different work situations in terms of information needs and in terms of required support to satisfy them.

Anna has just started working in the laboratory and she lacks experience in conducting her work. Since she needs some support, she looks for the experience accumulated by other scientists. She could ask her colleagues, but often they are either busy or out of the laboratory. She could also look in the notebooks of her colleagues, but there are two difficulties. The first one is to interpret the information found in the colleagues notebooks since they are organized according to their personal style and experience. The second one is to perform a "blind search" in the notebooks, without knowing a priori if the notebook may contain suggestions related to her current needs. To this aim, Anna needs an additional help since she could have not learned yet the conventions common to laboratory members for keeping their notebooks. However, due to her continuous presence in the lab, she is aware of the location of the materials she uses.

The situation is different for Paco, an experienced scientist, who is not always present in the laboratory but has a long experience of biological work. However, since he is not always present in the lab, he does not know where the materials he needs have been moved by other colleagues. Like in the case of Anna, it is sometimes problematic to ask other people also because he likes to attend the laboratory late in the evenings when nobody is there. To avoid wasting his time looking around in the shelves and fridges of the laboratory, Paco would like a support helping his search. On the other hand, due to his long working experience Paco could effectively help the other members of the laboratory by sharing the information reported in his notebook.

John attends the laboratory since a long time and usually he does not need any support. He is performing the same kinds of experiments since a long time, he knows where materials are, he has enough experience to perform his work, and he learned the conventions to compile the notebook. Due to his knowledge, he represents a useful source of information for the other colleagues. It is very important to recall here that although John does not need extra support to perform his work, nevertheless he actually provides information useful for the others. John's behavior, that might seem unrealistic (as it does not follow the logic of rewards [16]), is true in laboratory work, where the practice of compiling notebooks is part of the normal work. The reason is that laboratory work needs accountability, auditability and validability and these are achieved through the constant update of biologists' laboratory notebooks.

Summarizing the information issues in our scenario, scientists may need to get information about location of materials, action related to experiments, and sometimes about conventions for notebook compilation and use. There are different ways to collect this information either by asking colleagues or by consulting their notebooks. In the first case, problems may arise due to the different time schedules of the scientists in attending the laboratory. In the second case, either the searcher is not aware of the colleagues' notebooks contents or she is not able to interpret the collected information. The above considered problems are of the same nature: the need of collecting and sharing the valuable practices and the experiences of different scientists working in the lab. These problems may be solved in the same way by giving scientists the possibility to access this valuable knowledge and to get useful hints to complete successfully their experiments without waste of time and of personal budget.

4 Reading lab practices through Implicit Culture glasses

The concept of Implicit Culture [9] was introduced in Artificial Intelligence and in particular in the area of agents research. Implicit Culture is defined as the relation between two groups of agents such that the agents belonging to the second group behave consistently with the "culture" of the agents belonging to the first one. The technological counterpart of the Implicit Culture concept is a System for Implicit Culture Support (SICS) which is built upon collaborative filtering techniques. The system deals with arbitrary scenarios where users need suggestions concerning their work. In the scenario of biological work we considered, the system is used to support what is called in Nonaka's terms a *combination process* [13] to convert different bodies of explicit knowledge. In our proposal, this process concerns the reconfiguration of the knowledge the biologists embodied in their notebooks in terms of recurrent patterns of actions which the biologists performed during their work that the system discovers and makes available to biologists through the support it provides them.

Looking at the world with Implicit Culture glasses (and with SICS support) means that this world need to be observable, the actions need to be situated in a "computational" environment, along with the objects and the people involved in the actions.

Since our attention is focused on discovering the relation between actions performed by scientists in a laboratory and their information needs in order to support a biologist either to perform the next actions of an experiment or to better understand the causes determining the experiment failure, we adopt the same categorization we introduced above for the information needs, by distinguishing between *physical actions*, *scientific actions* and *social actions*.

According to the typology of considered action, different techniques may be used to observe and register the actions performed by biologists in the lab environment. Physical actions can be observed either tracking movements of objects in the lab with barcode like in the CyberFridge [17] or with surveillance technology using cameras along with object recognition techniques [18]. Scientific actions can be observed through the mediation of specific technology like Electronic Notebooks [19] or LIMS (Laboratory Informatics Management Systems) [20] which provide technologic frameworks where scientists can tune and control their experiments, but also of any technology supporting scientific production, like digital libraries and so on. Social actions are hardly observable from a technologic point of view except in particular cases where laboratory members are using a technology supporting collaboration and communication among them (like e-mail). However, usually biologists share the physical space of the lab and hence they rely on social mechanisms developed as a consequence of their co-presence.

5 The choice of notebooks as source of observation

There is a tradeoff between the ability to observe what happens in the laboratory and the richness of the support provided to scientists. Since notebooks are the main artifacts used by biologists to objectify work they perform in the laboratory, we focus on them in order to discover the relations among aspects that are relevant to scientists during their experiments. This information is then used to collect and make visible to biologists the practices encrusted in their notebooks. In this way, biologists are supported by the system during the conduction of their experiments maintaining a "peripheral awareness" [21] of their colleagues about performed actions and location of materials that would otherwise be invisible due to the implicit nature of their practices.

Protocols represent other kinds of artifacts that may be considered as biologists' reference for the practices related to the laboratory work. Biological protocols are written instructions that specify ingredients, equipments and sequence of steps for preparing and analyzing research materials [14]. Protocols may be either standardized procedures (in some cases sold by companies, in other cases provided as biological resources on the web [22] or collected in manuals) or hand-written procedures (codifying the experience accumulated by either a single scientist or inside a laboratory). Unlike [23], our focus is not on a technology designed to support workers in the execution of tasks associated to protocols. In fact as confirmed by observations of biologists' work in a laboratory and by related literature [14], usually actions reported in protocols are not slavishly executed by scientists but they are resources [24] which orient scientist in the completion of their work.

Instead, we focus on giving an indirect support to the scientists in deciding which actions to perform while "doing" the protocol [14]; actions may be suggested thanks to the visibility of laboratory practices. In fact, the latter may provide useful information about localization of materials, successful actions, and conventions to the scientists working in the same lab which may help scientists in performing actions.

We base our approach on the observation of notebooks and we argue this is feasible since notebooks are not totally unstructured due to the emerging of conventions related to their compilation as a consequence of the agreement among people working in the same laboratory. In the same vein Robinson et al in [21], base the reconstruction of processes related to work shifts in paper mill industries on diaries. The latter are artifacts and the reconstructed processes are artifact mediated processes.

The choice of notebooks as main source of observations implies requiring the scientists to do a technological shift from paper-based to electronic-based notebooks. In our opinion, this is not a critical issue for different reasons. First, experiments of PCR (like many others in a biological laboratory) are repeated several times introducing slight changes from one time to another and could therefore greatly take advantage from the use of electronic notebooks containing pre-defined working templates (as the one in Figure 2) and some kind of suggestions. Second, unlike the medical case where the shift from paper to electronic system becomes a critical issue due to the lack of a consolidated practice in compiling medical records as argued by Heath and Luff [25], in biological work the use of notebook is intrinsically related to the scientists' work practice. Accordingly, a wider set of Electronic Notebooks is proposed both commercially (see for instance The Gene Inspector by Textco Inc. [26] and E-Notebook by CambridgeSoft [27]) and in the research (see for instance [28]). A noticeable result of the research about electronic notebooks is represented by the Augmented laboratory notebook [29] which uses techniques of augmented reality to integrate the physical world of the biologists with a virtual one. Hence, the proposed notebook combines the flexibility of paper-based notebooks with the richness of computer supported information. In this way, the notebook is not disrupting the biologists' work practices due to its ability to correspond to its paper-based counterpart: for example it recognizes biologist's handwriting and adds the power of managing the biologists' information space at any place.

However, in our current research effort we did not choose any specific electronic notebook technology. From our point of view the choice of a specific technology is not a main concern. We advocate any technology allowing scientists to perform their work without disrupting their usual practices and leaving us the possibility to observe the biologists' transcriptions of the performed actions, including the related materials, machines used and people involved. Accordingly, we designed a first mock-up of the notebook interface focusing on the information which is possible to observe from transcriptions of biologists' work. This choice is motivated by the fact that we are interested in the information which is observable from the notebook and on how this information may be managed by the SICS technology to satisfy the biologists' information needs. In the next step, this initial mock-up will require further evaluation by the involved biologists and possibly a redesign by using other electronic notebook technologies as, for instance, the Augmented laboratory notebook described above.

6 Combining SICS with Electronic Notebooks

We consider two different scenarios of use of the proposed technology which are related to different phases of biologists' work. When a biologist is conducting an experiment, she has to perform actions in order to complete her work. When an experiment is completed it may result either in a success or in a failure according to the actions the biologists performed during the experiment. The support needed by the biologists changes according to these two different work situations. In the first case, during an experiment a biologist can not be able to perform the next action until a specific information need is satisfied. For instance before putting a reagent in the machine, the biologist has to know the reagent concentration or the temperature value to set. Or she needs to know how to proceed in her work, especially if she is a novice for the experiment she is running. The above situations require answers to the question “how can I proceed in completing the experiment?” and can be referred to in terms of “support for anticipation”. In the second case, the scenario considers a biologist who completed an experiment. If the experiment failed she may like to know “why did it happen?”. For instance she is interested in comparing her experience with the ones of other colleagues performing the same kind of experiment. What she needs is an “explanation of failure”.

In the first situation the way to fulfill the current scientist’s information need is to suggest a set of actions which, if performed, helps the biologist to answer the above questions. For instance, if the biologist finds a way know the needed concentration and the temperature value usually set for the procedure, she is able to continue her experiment. For what concerns the explanation of failure, the system helps the biologist in understanding the reasons of the failure. By suggesting the actions performed by colleagues in successful experiments of the same kind, the system allows the biologist to rethink her experiment by comparing the suggested actions with the ones she performed. Moreover, this learning process helps her to develop hypotheses about the failure of the experiment.

In this section we describe the architecture of a system that integrates Implicit Culture Support into an Electronic Notebooks (see Figure 3). The system takes into account the actions performed by the biologist in the physical space. Hence, the notebook of each biologist is the artifact which creates a link between the physical space where the biologists perform actions and the logical space where the system “observes” these actions. In order to simplify this flow of information, we designed a mock-up of the notebook interface which provides biologists with a semi-structure for keeping trace of the performed actions (see Figure 4). In the left part of the interface (see Figure 4), each action performed by the biologist is registered as a pair name of the action and set of parameters associated to it. Free text is then available for scientists in order to report personal annotations during the course of the experiment. Each action is then associated to a timestamp corresponding to the time of registration. In the right part of the interface (see Figure 4), the structure provided by the notebook makes it possible to register both the name of the scientist and the name of the experiments she runs. Moreover, since it is very important for a scientist to distinguish between successful and unsuccessful experiments, the notebook allows a biologist to register whether the experiment performed was successful or failed. The use of this semi-structure for registering scientific actions does not disrupt the biologists’ practices of registering the performed actions in their notebooks since the semi-structure is flexible enough and reflects the structure of the performed actions that we observed in the paper-based notebook (see Figure 1). In fact, the considered semi-structure takes into account that each experiment has a performer, a name identifying it on the notebook pages, the date of execution, and that a biologist evaluates the results of the performed experiment also in terms of either success or failure. Moreover, it considers that a relevant part of the information reported on the notebook is about the actions she performed in order to complete an experiment with the associated values.

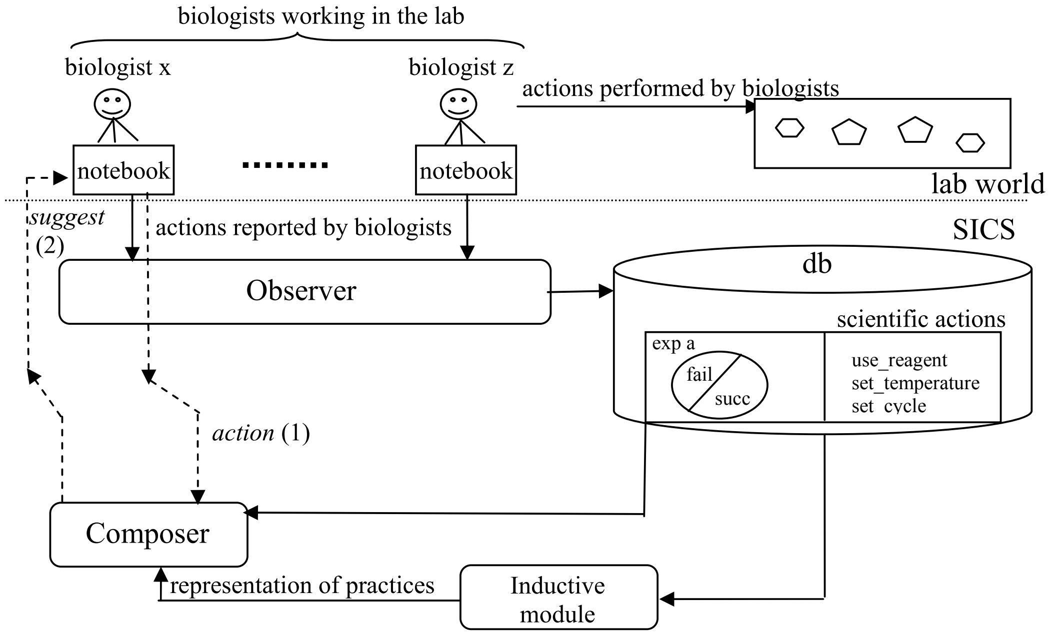


Figure 3: the SICS architecture in the laboratory scenario

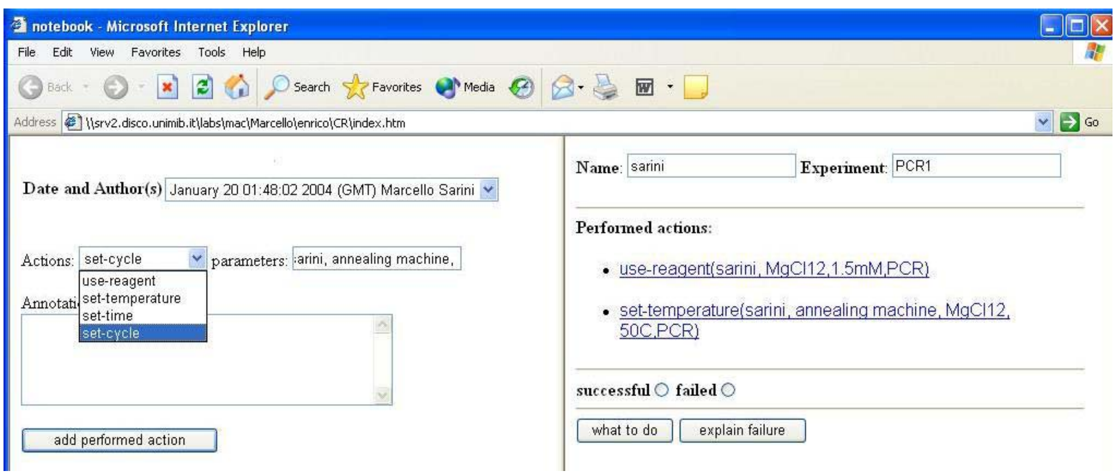


Figure 4: a mock-up of the notebook interface

As presented in [9], the architecture of a SICS is composed of three main modules (Figure 3): *Observer*, *Inductive*, and *Composer modules*.

6.1 The Observer module

The Observer observes the actions of biologists. We can distinguish two different typologies of actions to observe: (1) scientific actions performed by biologists during their work and that are written in their own notebooks; (2) actions that the biologists perform during the interaction with the electronic notebook, such as menu navigation or options choices. In this second case these actions could encompass the first ones as parameters. Since this second type strongly depends on the features of the designed interface of the electronic notebook we do not consider them at the present stage.

For what concerns scientific actions, we focused on a restricted set of meaningful actions which are a generalizations of the ones related to a PCR experiment (see for instance [22]). They refer to the percentage of reagent used in the current experiment; to the temperature the machine has been set during the reagent processing; to the time interval of reagent processing; to the number of cycle to be repeated in order to perform the experiment. This is in accordance with the concept of procedural abstraction advocated in the Labscape project (see [30]). In fact, according to the observations of work in a biological laboratory, the authors discovered that although laboratory work appears complex and tools and instruments are highly diverse, biologists perform only a few types of abstract operations, although in many different ways. More specifically, considering *x* a reference to the biologist performing an *experiment* by using a *machine* and a *reagent*, the actions are:

– *use-reagent(x, reagent, value, experiment)*, where *value* is the percentage used during the current experiment.
– *set-temperature(x, machine, reagent, value, experiment),* where *value* is the temperature set during the current experiment.
– *set-time(x, machine, reagent, value, experiment),* where *value* is the time interval the reagent is processed in the machine during the current experiment.
– *set-cycle(x, machine, reagent, value, experiment),* where *value* is the number of times the reagent is processed in the machine during the current experiment.

Then, the Observer module needs to collect together the scientific actions performed in a specific experiment and to distinguish between successful and unsuccessful instances of the same kind of experiment (see Figure 5). This is implemented by using a relational database where the scientific actions are stored and the needed views are created on the stored scientific actions. This way of organizing the information registered by biologists is then used by the system to relate the actions performed by biologists as described in the following.

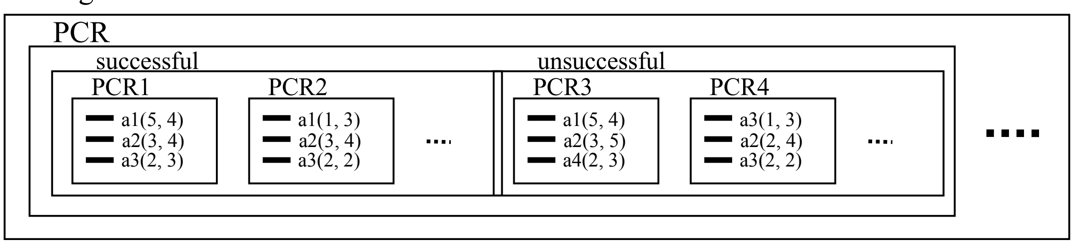


Figure 5: the views collecting scientific actions in the Observer database

6.2 The Inductive module

The Inductive module works on the scientific actions stored in the Observed module and discovers recurrent patterns which link them together. In this way the system supports the reconfiguration of the knowledge biologists externalized in their notebooks into patterns of actions in order to collect and make visible to biologists part of the implicit practices related to their work. The discovered patterns are expressed in terms of rules which have the following format:

$$A_1 \text{ and } A_2 \text{ and ... and } A_n \rightarrow C \quad (1)$$

where the antecedents of the rules are conjunctions of scientific actions and the consequent is a scientific action.

In order to discover patterns the Inductive module uses algorithms for mining association rules (like Apriori [31] in the WEKA implementation [32]).

We use two different strategies to discover rules which are of the first order. The choice of first-order rules is motivated by the fact that these rules have some parameters unspecified and hence they express more general patterns of association among scientific actions.

The first strategy uses the *Apriori* algorithm [31] to discover association rules involving actions which are grouped by the Observer module in the successful view. This is done by looking for actions whose parameters such as reagent, machine and temperature are frequently associated in the successful experiments. This algorithm is specifically designed to manage large amounts of data, typically for market analysis. The discovered association rules are propositional: this means that all their parameters have specified values. For instance consider that in successful experiments concerning PCR scientists consistently use the reagent MgCl2 with a concentration of 1.5mM, the annealing machine set with a temperature of 50°C running for 3 cycles. In this case, the Inductive module discovers and adds the following rule expressed where *x* is a generic scientist:

use-reagent(x, MgCl2, 1.5mM, PCR) and
set-temperature(x, annealing machine, MgCl2, 50 °C, PCR) →
set-cycle(x,annealing machine, MgCl2, 3, PCR) (2)

In order to generalize the rules mined by Apriori, our algorithm looks for actions with the same name and substitutes recurrent values of parameters with unspecified parameters. For instance consider that in successful experiments about PCR scientists used the reagent MgCl2 with a concentration ranging from 1.5mM to 3mM, with machines set with a temperature ranging from 50°C to 70°C and a variable number of cycles ranging fro 3 to 5, then the Inductive module discovers and adds the following rule:

use-reagent(x, MgCl2, c mM, PCR) and
set-temperature(x, m, MgCl2, t °C, PCR) →
set-cycle(x, m, MgCl2, N, PCR) (3)

where *c* represents the concentration of the reagent *MgCl2, m* the machine used by the biologist set with a temperature *t* and *N* represents the parametric value for the number of cycles; all these parametric values will be provided by the Composer module as described in the following.

The second strategy is used if the number of observed actions is limited; in this case it is possible to use mining algorithms such as *FOIL* [33] which directly obtains first-order rules like for instance:

use-reagent(x, MgCl2, c mM, PCR) →
set-temperature(x, annealing machine, MgCl2, t °C, PCR) (3bis)

where *c* represents a parametric value for the concentration which will be unified by the system with a value taken from an action performed by a biologist, while *t* representing a parametric value for the temperature will be provided by the Composer module as described in the following. The above rule simply says that biologists that uses a reagent for a PCR has to set the temperature of the annealing machine.

6.3 The Composer module

The Composer module is the core of the system, since it uses the information collected by both the Observer and the Inductive modules in order to propose a biologist the information which either satisfies her needs or helps her to understand why the experiment she performed failed. To illustrate how it happens (see dashed arrows in Figure 3) let us consider the flow of interaction inside the SICS system which starts from considering the need of a biologist up to providing the latter with the needed information. In step 1 an action performed by the biologist on the notebook triggers the Observer module to store it and starts the Composer module to select among the rules discovered by the Inductive module the ones whose left side is satisfied by the considered action. Then by using the selected rules, the Composer takes into account the previous biologists' experiences and proposes a suggestion to the biologist which should satisfy her information needs.
In particular this module consists of two main sub-modules:

- the Cultural Action Finder (CAF)
- the Scene Producer (SP)

6.3.1 The Cultural Action Finder (CAF)

Aim of the CAF module is to implement step 1 (see dashed arrow (1) in Figure 3) of the flow of interaction described above. Again, we have to consider the two different identified scenarios. During "support for anticipation" if a biologist wants to receive a suggestion from the system on the next action to perform, she requests it by asking through the notebook interface what to do next (see bottom right part of Figure 4). This request triggers the CAF to retrieve the last scientific action registered in the notebook of the biologist and hence stored by the Observer module. If the support provided by the system is related to "explanation of failure", when the biologist reports on her notebook interface the failure of an experiment, then if she wants to have some explanation for failure she may request it through the notebook interface (see bottom right part of Figure 4). This request triggers the CAF to iteratively retrieve all the scientific actions she performed in the failed experiment as registered in her notebook and stored by the Observer module.

Then, the CAF starts to consider the last action performed by the biologist as reported in her notebook. As a first step, the CAF has to select from the rules discovered by the Inductive module the one most suitable considering both the current action and the previous ones performed by the biologist during the same experiment. To do this selection the CAF tries to match the action executed by the involved biologist with the atoms belonging to the antecedents of the rules discovered by the Inductive module. This works by using a matching function which returns true when the names and parameters of the two actions match. Since it is possible that more rules may be selected, the algorithm sorts the rules by considering the number of atoms in the antecedent parts.

The algorithm then starts from the rules with the greatest number of atoms up to the ones with the lowest number and tries to select a rule which matches both the current scientific action and the actions the biologist performed in the same instance of the experiment. In this way the algorithm takes into account the current action, but also the previous history of the biologist who performed the considered actions and her experience in completing an experiment.

Once a rule is selected, the CAF looks whether there exists actions performed in successful instances of the same kind of experiment which satisfy the left part of the considered rule. This is done by looking in the Observer module. Hence, the actions belonging to the same successful experiment are collected together as single elements (a tuple) of a set if and only if each of them satisfies one of the atoms (representing an action) constituting the left part of the selected rule.

Then, the CAF returns as many actions as the cardinality of the set. Each action is the same reported in the right part of the selected rule but with different values of its variables. In fact, for each element of the set the variables of the action are unified with the corresponding values of the actions belonging to the considered element of the set. When there are no corresponding values for the parameters of the action, they are left unspecified.

In this way the system retrieves the past experience of the biologists dealing with a specific situation codified by a rule. The returned actions are called in Implicit Culture terms as *cultural actions* since in performing them a biologist behaves according to the culture of the others. In addition, these actions if performed by the biologist should satisfy her current information needs, in coherence both with the patterns previously discovered by the Inductive module and with the past actions performed by biologists.

Let us consider this situation: if during the analysis of the actions performed by a biologist in an unsuccessful experiment the action currently considered is *set-temperature(x,* *annealing machine, MgCl2, 50 °C, PCR)*, the CAF in the first step selects rule (3) since it is satisfied by the considered action. Then the CAF looks for successful past actions performed by other biologists satisfying the left part of the selected rule: for instance *use-reagent(x, MgCl2, 1.5mM, PCR)*. Hence, the CAF returns as cultural action the action *set-cycle(x, annealing machine, MgCl , N, PCR)* which is the result of the unification of the variables of the action in the right part of rule (3) with the corresponding values of the collected actions satisfying the left part of the rule. If the cultural action is specified in each of its parameters, for instance *set-cycle(x, annealing machine, MgCl, 3, PCR)*, the system can try to satisfy the biologist's current information needs by simply suggesting the considered action. Otherwise, if the cultural action has one or more of its parameters not specified, as *set-cycle(x, annealing machine, MgCl, N, PCR),* the role of the system is to suggest effective values for these parameters, in this case the number of cycles *N* to run the machine. In this last situation the job of the Scene Producer module is more complex since it has to find suitable values which better specify the selected cultural action among the actions stored by the Observer module.

6.3.2 The Scene Producer (SP)

The cultural actions determined by the CAF algorithm are then passed as input to the Scene Producer module which implements step 2 (see dashed arrow (2) in Figure 3) of the above described flow of interaction. In particular, for a given cultural action the SP generates on the biologist's notebook interface a suggestion such as it is maximized the probability that the biologist will perform next the cultural action proposed by the CAF. In case the cultural action is completely specified, the SP simply proposes it as suggestion to the biologists. In the other cases, the SP generates the suggestion for the biologist through an algorithm working in three steps.

1) *For each cultural action found by the CAF module, find the group of biologists which performed actions similar to the considered cultural action*. Again, computation of similarity is performed by looking in the Observer repository and by using a similarity function which works on the scientific actions. Here, scientific actions are similar to the considered cultural action if they belong to the same view, names match and the related parameters matches without considering the ones in the cultural action still with unspecified values. For instance if *use-reagent(x, MgCl2, c mM, PCR))* is the considered cultural action, *use-reagent (y, MgCl2, 3mM, PCR))* and *use-reagent(z, MgCl2, 4mM, PCR))* are similar to it if both of them belong to the same view which groups scientific actions performed in successful PCR experiments. Then *y* and *z* are references of the biologists returned by the sub-module in the first step of computation. At this point the SP has to use a strategy to select which action to consider as a relevant suggestion. This is done in the next step by looking for the biologists who are most similar to the one under concerns.

2) *Select among the biologists collected during step 1 a subset of biologists which are the most similar to the biologist who asked for suggestion*. Since biologists are working in the same laboratory and sharing common work practices, we could consider all of them similar. However, we want to consider here another form of similarity among biologists. In fact, in order to assess the similarity between the biologists we define a similarity function expressed as a sum of the similarities of the actions they performed stored by the Observer module. Namely, two biologists are similar if they reported in the notebooks similar actions. This form of similarity is used since we argue that similarity of biologists based on the performed actions is a reasonable way to relate the action to be suggested with the real needs of the considered biologist. A more fine grained similarity function which has still to be investigated could rank the biologists most similar to the one under concerns privileging the ones more experienced. In order to measure experience among biologists we consider for each of them the number of successful experiments versus the total number of completed experiments. In case all the identified scientific actions in the previous step were performed by the same biologist then the subset contains only the considered biologist which is obviously the most similar to the one under concerns. In this situation we provide as suggestion all the identified scientific actions.

3) *Once collected the actions similar to a cultural action performed by the most similar biologists, generate the suitable suggestions to the biologist in order to satisfy her needs.* This is done by generating a suitable clue on the biologist's notebook interface (for instance a pop up with the text of the suggestion). If the actions similar to the cultural action are too many it is possible to use a filtering technique that generalizes collaborative filtering [10]. Consider the case of explanation of failure, when considering an action *use-reagent(x, MgCl2, 2mM, PCR))* performed by the biologist *x* in a failed experiment. If the cultural action selected by the system is *set-cycle(x, annealing machine, MgCl2, 3, PCR))* then the related biologist *x* receiving it as a suggestion may compare the actions she performed in the experiment with the one proposed by the system. In this way the biologist is supported in rethinking the experiment she performed and in understanding, if possible, why it failed comparing her experience with the knowledge of her colleagues collected and proposed by the system.

7 Conclusions and future works

In this paper we proposed an approach and a technology to provide biologists with the support needed to complete their work. By observing the nature of laboratory work in the laboratories we identified situations where biologists have various information needs to be satisfied. Often biologists do not have the information they need at the right time and sometimes they lack the way of satisfying their needs. This because it may happen that colleagues which are a suitable source of information are busy or not present in the lab. Moreover, the scientific information they have access through common repository is hardly usable to satisfy a need in a specific situation.

We observed that a relevant work practice of biologists concerns the registration on their own notebooks of everything they do in a laboratory. Even though this information is freely available to people working in the same laboratory, often its interpretation is hard due to different personal organization of work and different levels of expertise. Hence, even if notebooks are artifacts where the tacit knowledge of biologists related to their experiences and practices is externalized, this knowledge is often not exploited by other biologists.

Adopting the Implicit Culture approach and the SICS technology our system tries to make the knowledge maintained by the notebooks more usable. In this way the SICS observing the transcriptions of the notebooks converts this knowledge in recurrent patterns of actions performed by biologists during experiments. By using these patterns expressed in terms of first-order rules the system is able to support the work of biologists by satisfying their needs. In particular, we identified two modalities of use of the system according to the biologists' needs: the support for anticipation and support for explanation of failure.

In the first case, it may happen that while performing an experiment a biologist may not be able to perform the next action until a specific information need is not satisfied. This is associated to an interactive use of the system: the system considers the last action performed by the biologist and suggests an action which if executed is coherent with the patterns of actions discovered by the system considering the successful experiments of the same kind. In this case, since the system suggests to the biologist an action to perform next, it may also be possible to observe on her notebooks if the suggested action is really performed by the biologist. Then this observation could be stored as an *accept()* action in the database. The use of this additional information of indirect feedback provided by biologists may be considered in a future extension of the system. In this way it may be possible to tune the support the system provides to the biologists taking into account the level of usefulness of the suggestion it proposed with respect to experience and preferences of the biologists working in the same lab.

In the second case, when the system is used to support the "explanation of failure" for an experiment, a biologist who performed an unsuccessful experiment would like to have some information in order to understand why it failed. In this situation the system works in "batch mode" that is considering all the actions performed by the biologist in the considered experiment. Then for each of them the system suggests the biologist the actions other colleagues performed in successful experiments of the same kind. Again, these actions are proposed by the system by considering the patterns of action discovered from the transcriptions of the notebooks.

There are several directions of future work. One direction is the investigation of techniques to discover more complicated patterns of actions, taking into account also sequence of actions. In fact our patterns, represented as first-order rules with conjunction of actions, do not express whether an action is performed before of another one.

A second direction will address other information needs. Though in this paper we focused on scientific needs, that is needs related to the scientific actions performed by biologists during an experiment and registered in their notebooks, we argue that the system can be extended to support biologists to satisfy other needs. In particular we want to focus our future work on physical needs: that is, to have some suggestions about the location of a reagent by considering the transcriptions of the notebooks in the database of scientific actions. In fact, since this database stores scientific actions, it can answer queries like "which is the last action performed by using a specific reagent?". Hence, the answer contains references either to the scientist who performed the action or to the machine used during the considered action. In this way, the system integrating the notebooks with the database of scientific actions indirectly provides the biologists with hints about the location of a reagent according to the scientific actions performed. In addition, since the system is able to observe actions performed by biologists on their notebooks, it is possible to register in the Observer module the queries performed by biologists to request for a reagent. Then these actions may be used by the Composer module to suggest a biologist, querying for a reagent, to contact the biologists who asked for it previously. This represents an extension of the system towards the suggestion of social actions to satisfy biologists' information needs.

Acknowledgements

Many thanks to Carla Simone for strong support, encouragements, and suggestions; to Michele Telaro and Jasminko Novak for useful discussions.

References

[1] K. Schmidt and C. Simone, "Coordination Mechanisms: Towards a conceptual foundation for CSCW systems design," *Computer Supported Cooperative Work, The Journal of Collaborative Computing*, vol. 5, pp. 155-200, 1996.

[2] M. Klein (editor), "Special Issue on Adaptive Workflow Systems," *Computer Supported Cooperative Work, The Journal of Collaborative Computing*, vol. 9, 2000.

[3] K. Schmidt (eds), C. Heath, and T. Rodden, "Special Issue on Awareness in CSCW," *Computer Supported Cooperative Work, The Journal of Collaborative Computing*, vol. 11, 2002.

[4] G. Mark, "Conventions and Commitments in Distributed CSCW Groups," *Computer Supported Cooperative Work, The Journal of Collaborative Computing*, vol. 11, pp. 349-387, 2002.

[5] L. Bannon and S. Bodker, "Constructing Common Information Space," ECSCW'97, Lancaster (UK), 1997.

[6] M. Sarini and C. Simone, "Recursive Articulation work in Ariadne: the alignment of meanings," COOP2002, Saint Raphael, France, 2002.

[7] U. M. Borghoff and J. H. Schlichter, *Computer-Supported Cooperative Work: Introduction to Distributed Applications*: Springer Verlag, 2000.
[8] A. D. Marwick, "Knowledge management technology," *IBM SYSTEMS JOURNAL special issue on Knowledge Management*, vol. 40, 2001.
[9] E. Blanzieri and P. Giorgini, "From Collaborating Filtering to Implicit Culture: a general agent-based framework," Workshop on Agent-Based Recommender Systems (WARS) in Autonomous Agents (Agents2000), Barcelona, Spain, 2000.
[10] E. Blanzieri, P. Giorgini, P. Massa, and S. Recla, "Information access in Implicit Culture framework," Tenth ACM International Conference on Information and Knowledge Management (CIKM 2001), Atlanta, Georgia, 2001.
[11] E. Blanzieri, P. Giorgini, F. Giunchiglia, and C. Zanoni, "Personal Agents for Implicit Culture Support," AAAI Spring Symposium on Agent-Mediated Knowledge Management (AMKM-03), Stanford CA, 2003.
[12] P. Resnick and H. R. Varian, "Recommender Systems," *Communication of the ACM*, vol. 40, pp. 56-58, 1997.
[13] I. Nonaka and H. Takeuchi, *The Knowledge Creating Company*. Oxford, UK: Oxford University Press, 1995.
[14] M. Lynch, "Protocols, practices, and the reproduction of technique in molecular biology," *British Journal of Sociology*, vol. 53, pp. 203-220, 2002.
[15] Lab Notebook Description in BioSci 1510, at http://www.pitt.edu/~biohome/Dept/docs/syllabi/1510_031.pdf Fall Term 2002.
[16] J. Grudin, "Why CSCW Applications Fail: Problems In The Design And Evalutation Of Organizational Interfaces," Proceedings CSCW'88, Portland, Oregon, USA, 1988.
[17] J. Mankoff, J. Somers, and G. D. Abowd, "Bringing people and places together with Dual Augmentation," Proceedings of the 1998 Spring AAAI Symposium on Intelligent Environments, 1998.
[18] G. Andolfi, M. Aste, M. Boninsegna, R. Cattoni, A. Potrich, and B. Caprile, "The Advanced Visual Monitoring Project at IRST," in *Advanced Video-Based Surveillance Systems*: Kluver Academic Publisher, 1998.
[19] Collaborative Electronic Notebook System Association (CENSA), at www.censa.org.
[20] LIMS Guide on Online Journal Scientific Computing and Instrumentation, at http://www.scimag.com/Scripts/lims.asp?SECTYPE=LIMS 2003.
[21] M. Kovalainen, M. Robinson, and E. Auramaki, "Diaries at Work," Proceedings CSCW'98, Seattle Washington, 1998.
[22] bioProtocol. a Bio Online Site, at http://bioprotocol.bio.com 2003.
[23] B. Katzenberg, F. Pickard, and J. McDermott, "Computer Support for Clinical Practice: Embedding and Evolving Protocols of Care," Proceedings CSCW'96, Cambridge MA, 1996.
[24] L. A. Suchman, *Plans and situated actions. The problem of human-machine communication*. Cambridge: Cambridge University Press, 1987.
[25] C. Heath and P. Luff, "Documents and Professional Practice: 'bad' organisational reasons for 'good' clinical records," Proceedings CSCW'96, Cambridge MA, 1996.
[26] Gene Inspector, at http://www.textco.com/products/gene_inspector.html.
[27] P. Minoofar, "E-Notebook A Digital Alternative to the Mighty Lab Notebook," *ChemNews.com; http://chemnews.cambridgesoft.com/art.cfm?S=275*.
[28] The ORNL Electronic Notebook Project, at http://www.csm.ornl.gov/~geist/java/applets/enote/.
[29] W. E. Mackay, G. Pothier, C. Letondal, K. Boegh, and H. E. Sorensen, "The missing link: augmenting biology laboratory notebooks," Proceedings of UIST'02, Paris, France, 2002.
[30] L. Arnstein, C. Hung, Q. H. Zhou, G. Borriello, S. Consolvo, and J. Su, "Labscape: A Smart Environment for the Cell Biology Laboratory," *IEEE Pervasive Computing*, vol. 1, pp. 13-21, 2002.
[31] A. Rakesh and S. Ramakrishnan, "Fast Algorithms for Mining Association Rules," Proc. 20th Int. Conf. Very Large Data Bases (VLDB), 1994.
[32] Weka 3 - Data Mining with Open Source Machine Learning Software in Java, at http://www.cs.waikato.ac.nz/ml/weka/.
[33] J. R. Quinlan, "Determinate Literals in Inductive Logic Programming," Proceedings 12th International Joint Conference on Artificial Intelligence, 1991.